Descrizioni archivistiche e web semantico: un connubio possibile?

Salvatore Vassallo

Introduzione

Il presente articolo è incentrato sulla valutazione delle possibilità, dei vantaggi e delle problematiche nell'esprimere le descrizione archivistiche direttamente nel web semantico. Una riflessione in questo ambito non ha tentativi precedenti almeno per quanto concerne lo specifico del mondo archivistico, mentre recentemente una simile riflessione è stata condotta da Martha Yee, con una ricerca volta a indagare se (e con quali caratteristiche) i dati bibliografici possano sopravvivere direttamente nel web semantico (Yee 2009). Per verificare se le descrizioni archivistiche possano sopravvivere direttamente nel web semantico è necessario valutare in che modo rendere ogni singola regola e esigenza prevista dagli standard archivistici, traducendoli negli specifici costrutti della tecnologia scelta (nel caso in oggetto le *Topic Maps*, standard International Organization for Standardization (ISO) 13250). Una volta assicurato il rispetto delle regole previste dagli standard è, però, necessario valutare quali conseguenze e quali scenari di utilizzo possano aprirsi nella diffusione di informazione così codificata.

JLIS.it. Vol. 1, n. 1 (Giugno/June 2010), p. 169–193.
DOI: 10.4403/jlis.it-27

Gli strumenti usati: le *Topic Maps*

Topic Maps e RDF

Martha Yee, nelle sue ricerche e nel saggio citato in precedenza, costruisce un modello basato su Resource Description Framework (RDF) per esprimere e vincolare le regole di catalogazione, guidando, in questo modo, la pubblicazione dei dati catalografici direttamente nel web semantico. Le riflessioni alla base di questo articolo saranno rivolte invece all'espressione di descrizioni archivistiche utilizzando le *Topic Maps*, una tecnologia codificata dalla famiglia di standard ISO 13250. Lo standard ISO 13250 è infatti uno standard a più livelli (parti) che definisce la tecnologia delle *Topic Maps*, in particolare si compone di:

- 13250-1 *Topic Maps* - Overview and Basic Concepts: un'introduzione discorsiva alle *Topic Maps*
- 13250-2 *Topic Maps* Data Model: il modello di dati delle *Topic Maps*
- 13250-3 *Topic Maps* - XML Syntax: una sintassi XML per la serializzazione di *Topic Maps*
- 13250-4 *Topic Maps* - Canonicalization Canonicalization XML *Topic Maps* (CXTM): è anch'essa una sintassi XML, a differenza della precedente non orientata all'interscambio e alla comunicazione delle informazioni, ma finalizzata alla verifica della conformità dei software nel supportare i diversi linguaggi di serializzazione
- 13250-5 *Topic Maps* Reference Model: un modello di riferimento più astratto del TMDM che, pur con un'indubbia utilità, identifica in parte arbitrariamente costrutti come *topics*, *occurrences*, *names* e *associations*.

- 13250-6 *Topic Maps* - Compact Syntax: una sintassi per serializzare *Topic Maps* meno verbosa e prolissa di XTM
- 13250-7 *Topic Maps* - Graphical Notation: nasce per rispondere all'esigenza di una notazione grafica utile sia per documentare e rappresentare in esempi *Topic Maps*,sia per descrivere l'ontologia e il modello

A questi vanno aggiunti gli standard:

- ISO 18048 *Topic Map* Query Language: un linguaggio di interrogazione specifico per questa tecnologia
- ISO 19756 *Topic Maps* Constraint Language: un linguaggio per esprimere schemi di validazione
- ISO 29111 *Topic Maps* - Expressing Dublin Core Metadata using *Topic Maps*: una mappatura volta a definire come esprimere metadata Dublin Core in *Topic Maps*.

Entrambe le tecnologie, *Topic Maps* e RDF, sono cardini nella lunga strada verso il web semantico e condividono aspetti e potenzialità di utilizzo, pur prefigurando scenari di utilizzo differenti (Pepper 2008b). In passato, fin dai primi anni del 2000 agli albori di entrambe le tecnologie, si è spesso riflettuto sulla necessità e profitto nell'esistenza di due distinte tecnologie e le prime valutazioni a riguardo sono volte a stabilire i legami esistenti fra le *Topic Maps* e gli altri modelli, schemi e sintassi legati al web semantico e a vagliare la possibilità di fusione fra questi linguaggi (Garshol 2001; Pepper 2002). Tuttavia, lo stesso Garshol (2003) ha in seguito concluso che una vera e propria fusione degli standard (in particolare delle *Topic Maps* e di RDF) non sembra essere possibile né desiderabile e le discussioni si sono orientate verso vocabolari di mapping che permettano un interscambio di dati tra le due tecnologie e i differenti formati (Pepper *et al.* 2006a,b; Presutti *et al.* 2005).

Aldilà delle differenze implementative, degli strumenti di sviluppo e dei differenti scenari di utilizzo delle due tecnologie, la scelta di utilizzare le *Topic Maps* come strumento per esprimere nel web semantico descrizioni archivistiche conformi agli standard descrittivi è dettata dalla possibilità di utilizzare il *Topic Maps* Constraint Language (TMCL), standard ISO 19756 giunto allo stadio di Final Committee Draft (FCD) (ISO/IEC FCD 19756 2010), per esprimere vincoli e obbligatorietà nelle descrizioni. Il TMCL è dunque un linguaggio vincolato per creare schemi di validazione per le *Topic Maps*[1] Utilizzando il TMCL è possibile esprimere obbligatorietà e ripetibilità degli elementi e dei costrutti previsti all'interno del *Topic Maps* Data Model (TMDM) (ISO/IEC 13250-2:2006 2006) o vincolare (anche tramite *regular expression*) il valore che questi possono assumere. Per quanto concerne l'aspetto della validazione, questo linguaggio vincolato per Topic Maps risulta superiore all'accoppiata RDF e Ontology Web Language (OWL), seppur anche quest'ultimi possano esprimere diversi vincoli, ma rimanendo in ogni caso strumenti orientati e strutturati per l'inferenza e per una "semantica del ragionamento" rispetto a una validazione vincolante (Garshol 2006). Per perseguire l'obiettivo di verificare se e in che misura le regole di descrizione archivistica possano essere espresse nel semantic web, risulta maggiormente pratico l'utilizzo di una tecnologia con specifici standard preposti alla validazione.

Breve introduzione alle *Topic Maps*

Le *Topic Maps* sono una tecnologia per la rappresentazione e l'interscambio della conoscenza. Nonostante il nome della tecnologia e le differenti traduzioni ormai stratificate e diffusesi nella letteratura

[1]Sostanzialmente è possibile paragonare il TMCL a ciò che rappresentano XML Schema e Document Type Definition (DTD) nel caso dei documenti XML.

italiana sull'argomento (Meschini 2005; Vassallo 2007; Vivanet 2007; Weston 2002), le *Topic Maps* non sono né mappe di navigazione,[2] né mappe grafiche, [3] né mappe mentali o concettuali (Garrido e Tramullas 2004). Infine, ciò è essenziale nel prosieguo e nello sviluppo della discussione, non devono essere identificate necessariamente con eXtensible Markup Language (XML): XML *Topic Maps* (XTM) è solo uno dei possibili formati di serializzazione, ma una topic map può essere immagazzinata anche all'interno di un database relazionale o di altri lignuaggi di serializzazione (Garshol 2005). La flessibilità di questa tecnologia, evolutasi dal contesto di creazione e fusione di indici analitici di manuali tecnici entro cui era nata con il nome di Navigation Maps (Pepper 1999), offre alcuni strumenti che possono renderla appetibile anche per la gestione e fusione di metadati descrittivi differenti (Garshol 2004). Il modello di dati delle *Topic Maps* definisce infatti alcuni elementi (costrutti nel lessico delle *Topic Maps*) che combinati fra di loro permettono di esprimere e codificare una gamma illimitata di descrizioni e relazioni e che, per tornare allo scopo originario della tecnologia, permettono di gestire agevolmente indici multipli fondendoli e disambiguando i differenti termini. Senza illustrare nel dettaglio i singoli costrutti che costituiscono questa tecnologia, si possono evidenziare quattro elementi fondamentali che è necessario prendere in esame per sottolineare il grado di flessibilità che è possibile raggiungere con applicazioni basate su *Topic Maps*.

---

[2]Tuttavia è possibile utilizzare le *Topic Maps* per emulare mappe di navigazioni, come avviene per il software xSiteable http://xsiteable.sourceforge.net/.

[3]Per visualizzare Topic Maps è possibile utilizzare librerie grafiche come Hypergraph http://hypergraph.sourceforge.net/, Touchgraph http://sourceforge.net/projects/touchgraph/ o il modulo Vizigator all'interno della suite Ontopia http://code.google.com/p/ontopia/, per visualizzarle sotto forma di grafo, utile ad esempio per navigare ontologie.

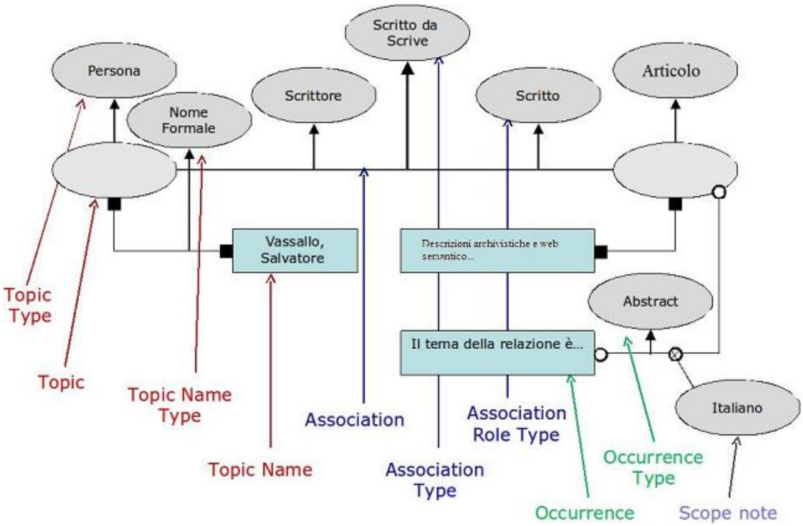


Figura 1: Esempio di *Topic Maps* espresso nella notazione grafica GTMalpha, una proposta di schema grafico per la produzioni di esempi di istanze di topic map (Thomas *et al.* 2008)

Topics

Il *topic* è la rappresentazione univoca di un qualunque ''soggetto''', intendendo per soggetto qualunque elemento del discorso su cui l'autore della *topic map* intenda parlare. *Topics* e *subjects* sono dunque gli ideali componenti di un triangolo semiotico: rappresentano rispettivamente il simbolo e il referente (Eco 1973; Ogden e Richards 1923). Tra il *topic* e il *subject* esiste una relazione uno a uno, in cui ogni soggetto viene rappresentato da un solo topic e ogni topic può rappresentare un unico soggetto. Questo principio (detto *collocation objective*, ''collocazione oggettiva") assicura che tutto ciò che si conosce (all'interno di un dato sistema) su un determinato soggetto sia accessibile univocamente da un punto d'accesso, dato che quel soggetto è rappresentato da solo un *topic* (Sigel 2003).

Ad esempio sono *topics* "JLIS", "Salvatore Vassallo", "Descrizioni archivistiche e web semantico un connubio possibile?" etc. Il *topic* può essere caratterizzato tipologicamente, topic type (tipo di argomento), sia per favorire l'aumento delle informazioni (si riconduce un'istanza a una determinata classe), sia per la risoluzione di omonimie: si pensi a Vassallo (*topic type*: persona) rispetto a Vassallo (*topic type*: feudatario) e a tutti gli stratagemmi anche grafici utilizzati per disambiguare i due termini nella creazione di indici (si tratta ancora una volta di dover gestire stesse stringhe di testo che identifichino soggetti differenti). Continuando l'esempio precedente topic types potrebbero essere "Rivista", "Persona", "Articolo"; è essenziale considerare che i topic types saranno a loro volta topics: sarà necessario, ad esempio, dichiarare esplicitamente[4] "Rivista",

[4]In realtà la nuova proposta di formato di scambio XML, XTM 2.1, permette di indicare il tipo di argomento anche facendo riferimento direttamente a un *subject identifier* o a un *subject locator*, ossia a un Uniform Resource Identifier (URI) con funzione di identificativo, senza doverlo necessariamente includere formalmente come topic all'interno della topic map.

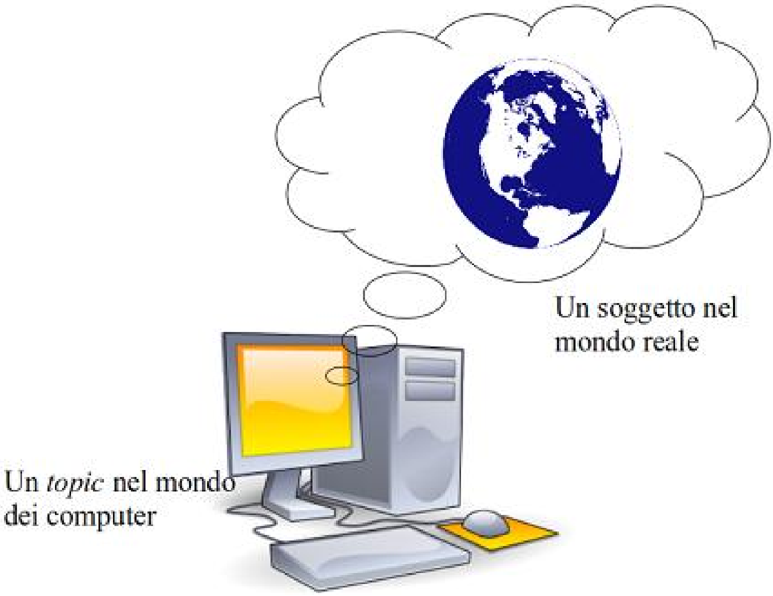


Figura 2: I topics sono surrogati (o proxies) nel mondo dei computer per gli altrimenti ineffabili soggetti di cui si intende discutere.

"Persona", "Articolo", come *topics* nella propria *topic map* per poterli associare come tipi delle istanze del nostro sistema. Ogni topic può avere diversi nomi, anche questi caratterizzabili tipologicamente attraverso il *topic name type* (il tipo di nome dell'argomento). Nell'esempio di 1 si indica, ad esempio, un *topic* di tipo persona, che ha un nome di tipo "Nome formale" e topic name "Vassallo, Salvatore".

Associations

Un'associazione è un costrutto delle *Topic Maps* per rappresentare una relazione fra due o più topics (definiti *role players*, ''attori dell'associazione').

Tipicamente le relazioni (e dunque la loro rappresentazione in un'associazione) sono binarie, ma è possibile instaurare associazione tra più topics (3-ary o n-ary se fra un numero illimitato di soggetti) come ad esempio nel caso "Il complesso archivistico X, prodotto da Y è conservato presso Z"[5] (Ahmed 2003; Bry *et al.* 2006; Pepper 2000). L'associazione viene definita da un association type (tipo di associazione) che permette di esprimere una semantica della relazione indicandone la tipologia. Si noti che anche i tipi di associazione, come la maggior parte dei costrutti delle *Topic Maps*, sono a loro volta *topics* (questa considerazione risulterà fondamentale quando, nel corso della ricerca, si cercheranno di evidenziare le caratteristiche di flessibilità di questa tecnologia). L'associazione inoltre può essere orientata in modo da determinare l'esatto ruolo che il topic assume nell'associazione (member role, ruolo dell'attore) con la possibilità, dal punto di vista dell'interfaccia utente, di determinare il verso delle associazioni per evitare paradossi come "Descrizioni archivistiche e web semantico scrive Salvatore Vassallo", ma anche per chiarire

[5] L'esempio ricalca una delle relazioni previste nella sezione dei collegamenti fra descrizioni dello standard archivistico International Standard for Describing Institutions with Archival Holdings (ISDIAH).

il ruolo dei singoli topics in una relazione del tipo ″Y è padre di X″ o in una relazione gerarchica di intero/parte. Nell'esempio grafico proposto in apertura di sezione, si rappresenta un'associazione di tipo ″Scrive/È scritto da″ tra il topic ″Vassallo, Salvatore″ che recita il ruolo di ″scrittore″ e il *topic* ″Descrizioni archivistiche e web semantico″ che ha il ruolo di ″scritto″[6].

Occurrences

Le occorrenze possono essere viste come associazioni interne al topic stesso: sono, a tutti gli effetti, una rappresentazione di una relazione tra un soggetto e una risorsa informativa (ovvero la rappresentazione di qualunque sequenza di byte, eventualmente recuperabile in rete). Questa considerazione deve essere valutata con attenzione nella fase di modellazione del proprio sistema: in alcuni casi l'utilizzo di occorrenze per esprimere quelle che in realtà sono associazioni binarie può essere visto come una forzatura (). Le occorrenze possono essere intese come i campi descrittivi di un topic e anche queste sono caratterizzate da uno specifico occurrence type che ne determini la natura: nell'esempio proposto è presente un'occorrenza, all'interno del topic ″Descrizioni archivistiche e web semantico″ di tipo ″abstract″.

Scope

Il modello delle *topic map* permette di fornire tre tipi di asserzioni (definite complessivamente come le caratteristiche del *topic*, o *topic characteristics*) su un *topic*: il suo nome, le sue occorrenze e

[6] In realtà il tipo di associazione, a propria volta un topic, potrebbe avere due differenti topic name (″scrive″ e ″è scritto da″) che potrebbero essere utilizzati dall'interfaccia in fase di visualizzazione a seconda del ″verso″ della relazione e quindi del ruolo del topic coinvolto.

le associazioni in cui è coinvolto; la specificazione di una caratteristica avviene, tuttavia, esclusivamente all'interno di uno specifico contesto (il problema del contesto del resto riveste un ruolo chiave nel mondo archivistico e nella descrizione della documentazione). *Scope* ("àmbito") è appunto un costrutto delle *Topic Maps* definito nel modello dei dati con la funzione di permettere di limitare l'ambito di validità di un'asserzione, in questo senso una dichiarazione priva di un ambito specificato formalmente è definita '*unconstrained*' ovvero illimitata (non dipendente da un determinato e specifico contesto). Nel grafico di Figura 1, ad esempio, si limita l'occorrenza di tipo abstract qualificandola linguisticamente. Il costrutto di scope può essere dunque utilizzato per limitare un'asserzione da diversi punti di vista: questo costrutto rappresenta un elemento fondamentale usato in casi che spaziano dal multilinguismo all'espressione dei limiti temporali di un'asserzione. Nel passato soltanto Marc De Graauw, Steve Pepper e Geir Ove Grønmo hanno evidenziato alcuni spunti di discussione sull'utilizzo del costrutto scope (De Graauw 2002a,b; Groenmo e Pepper 2001), mentre recentemente Lars Marius Garshol ha sollecitato una riflessione sulla definizione della semantica di scope, sulla sua traduzione pratica, sugli effetti sulle interfacce, sulle possibilità offerte di filtraggio e, dal punto di vista informatico, sull'impatto sui linguaggi di interrogazione (Garshol 2008). In estrema sintesi, sulla scorta della riflessione di Garshol, è possibile evidenziare, senza pretese di esaustività, alcuni settori di applicabilità del costrutto di scope come:

- **multilinguismo**: un nome o un'occorrenza possono essere limitati dal punto di vista linguistico o segnalando una variante dialettale o regionale. Questa possibilità, nel caso, dei nomi, è amplificata dall'opportunità di specificare diversi topic name type (tipo di nome dell'argomento) e variant (variante). Dal lato dell' interfaccia ciò si dovrebbe tradurre con la navigazione

di ontologie multilingue, passando da un linguaggio all'altro, escludendo specifiche lingue etc;

- **provenienza**: è possibile indicare la fonte di una determinata dichiarazione; ciò riveste particolare importanza, ad esempio, nel caso di datazioni di un documento, laddove le informazioni sulla data (cronologica e topica) siano provenienti da fonti differenti (eventualmente in contrasto fra di loro);
- **autorità – opinione**: è possibile segnalare che un'asserzione è valida in accordo con una determinata autorità o è una opinione di una determinata persona. Questo da un lato è essenziale per la costruzione di authority files, permettendo, ad esempio, di indicare quali regole di catalogazione o indicizzazione determini l'intestazione. L'utilizzo di scope per limitare un'asserzione a un'opinione personale, apre la possibilità a modifiche della struttura e dell'ontologia personali, valide esclusivamente per un singolo o per un gruppo - *cluster* - di utenti con simili interessi (Vassallo e Weston 2007);
- **tempo**: i limiti temporali di validità di un'asserzione sono essenziali per esprimere l'evoluzione diacronica di un termine, di una denominazione, di un'associazione; ad esempio in questo modo è possibile limitare cronologicamente (per il periodo di durata della collaborazione) l'associazione fra una persona e un'organizzazione etc.;
- **pubblico**: la possibilità di indicare il destinatario di un'asserzione è di estrema importanza, soprattutto laddove si voglia strutturare descrizioni pensate per target differenti. Dal punto di vista della resa per l'utente questo può avere diverse implicazioni: da filtri e esclusione di termini e istanze a seconda del pubblico fino a proporre descrizioni e occorrenze differenti a seconda dei diversi utenti.

Esprimere descrizioni archivistiche in *Topic Maps*

L'obiettivo di verificare i vantaggi e le criticità nell'immettere e gestire le descrizioni archivistiche direttamente nel web semantico richiede una solida base che confermi che qualunque esigenza o regola di descrizione prevista dagli standard archivistici sia gestibile con sufficiente granularità. Pertanto è necessario mappare gli standard descrittivi archivistici, nello specifico General International Standard Archival Description (ISAD(G)), International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF)), International Standard for Describing Institutions with Archival Holdings (ISDIAH) e International Standard for Describing Functions (ISDF), con il modello dei dati delle *Topic Maps* in maniera tale da fornire indicazioni su come esprimere le differenti esigenze evidenziate dagli standard archivistici con i costrutti delle *Topic Maps*. All'interno del presente articolo non si potrà fornire una disamina dettagliata di ogni singolo elemento, ma ci si limiterà a tratteggiare i capisaldi per una simile ricerca e a fornire alcuni esempi, volti a dimostrare la flessibilità dello strumento scelto. Come già accennato in precedenza una simile ricerca non ha in letteratura omologhi. I punti di partenza del lavoro saranno dunque:

- il citato studio di Martha Yee che, pur utilizzando una tecnologia differente (RDF) e rivolgendosi a un altro mondo (i dati bibliografici), evidenzia i passaggi metodologicamente necessari in un approccio di questo genere (da questo punto di vista, la mappatura tra standard archivistici e il data model delle *Topic Maps* può essere vista come un omologo del modello RDF[7] che Martha Yee costruisce per esprimere le regole di

[7] Si veda http://myee.bol.ucla.edu/rdfmodelintro.html.

catalogazione);

- le ricerche del gruppo di lavoro del professor Sam Oh, incentrate sulla possibilità di esprimere le relazioni previste dal rapporto Functional Requirements for Bibliographic Records (FRBR) in *Topic Maps* e di convertire direttamente record MARCXML (Lee, Jeon e Han 2006; Oh 2008);

- il tentativo di definire linee guida per esprimere metadati Dublin Core in *Topic Maps* poi confluito nella definizione dello standard ISO/IEC CD 29111 2009.

In particolare il terzo caso riveste grande importanza per il presente lavoro di ricerca in quanto traccia chiaramente i requisiti e le difficoltà di simili approcci (Maicher 2008; Pepper 2008a). Nel caso del Dublin Core si è trattato di definire le modalità e i costrutti idonei per gestire il set base di 15 elementi[8] (in seguito esteso a tutti gli elementi previsti da DCTERMS[9]) spaziando da soluzioni minimaliste (una serie di coppie chiave/valore rappresentate in occorrenze) a soluzioni, poi adottate, maggiormente strutturate con la definizione di associazioni per rappresentare le varie relazioni implicite a un record Dublin Core (autore, lingua, editore etc.). L'importanza, sottolineata in questo articolo in più occasioni, di utilizzare schemi di validazione per quanto sommaria è stata più volte ribadita e calendarizzata all'interno del gruppo di lavoro sullo standard ISO 29111.[10]

[8] Si veda http://myee.bol.ucla.edu/rdfmodelintro.html>.

[9] DCMI Metadata Terms http://dublincore.org/documents/dcmi-terms/>.

[10] Si veda Support #1342: Investigate if we should use TMCL to define constraints on DC all'indirizzo http://projects.topicmapslab.de/issues/1342.

Un esempio di mappatura fra il modello di dati delle Topic Maps e gli standard archivistici

Il seguente (Figura 3) è un semplice esempio di schema, presupposto necessario per mappare i diversi elementi previsti dagli standard archivistici nel data model delle *Topic Maps*. Si tratta dell'area dell'identificazione prevista da ISAAR(CPF).

Nello specifico si definisce innanzitutto il tipo di *topic* "Agente", che è un abstract ovvero non potrà avere istanze direttamente collegate, ma queste saranno legate alle sue sottoclassi "Ente", "Persona", "Famiglia". Le forme autorizzate del nome, le forme parallele del nome, le forme del nome autorizzate secondo altre regole e le altre forme del nome, saranno tutte tipi di nomi del topic, che possono essere ripetute a piacere con l'obbligo che sia presente almeno un topic name di tipo forma autorizzata del nome. Si stabilisce inoltre che questi *topic names* possano essere qualificati da scope notes di tipo data, periodo, qualificazione, lingua (nel caso delle forme parallele), norma (nel caso delle forme autorizzate secondo altra regola). Infine si definisce che i topics di tipo "Ente" possano avere un'occorrenza di tipo "Codice identificativo dell'ente", che sia univoca (cioè che assuma sempre valore diverso), limitando il tipo di dato accettato per il valore dell'occorrenza a NMTOKEN[11]. Tutti questi vincoli possono essere poi codificati e espressi in una sintassi CTM (un linguaggio di serializzazione compatto per *Topic Maps*, ad esempio:

```
isaar-tm:agente isa tmcl:topic-type;
- 'Agente';
is-astract();
```

[11] http://www.w3.org/TR/xmlschema-2/#NMTOKEN.

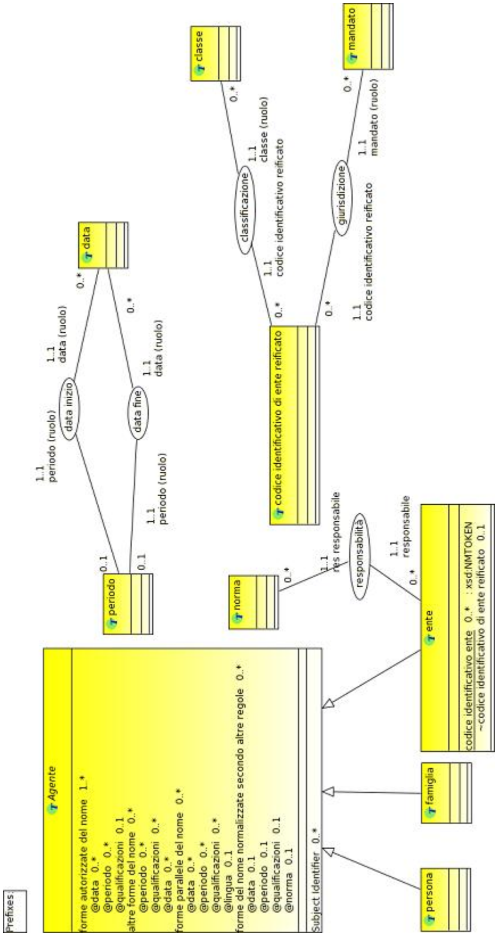


Figura 3: Esempio di modellazione dell'area dell'identificazione delle ISAAR(CPF), espresso in una notazione grafica simil Unified Modeling Language (UML) utilizzando Onotoa, un software per la creazione grafica di schemi TMCL.

```
has-name(isaar-tm:forma-autorizzata, 1, *).
isaar-tm:forma-autorizzata isa tmdm:topic-name-type;
- 'Forma autorizzata del nome
http://gilgamesh.unipv.it/psi/isaar/#isaar512>.
isaar-tm:persona isa tmcl:topic-type;
- 'Persona';
ako isaar-tm:agente.
isaar-tm:famiglia isa tmcl:topic-type;
- 'Famiglia';
ako isaar-tm:agente.
isaar-tm:ente isa tmcl:topic-type;
- 'Ente';
ako isaar-tm:agente;
has-occurrence(isaar-tm:codice-identificativo-ente, 0, *)
isaar-tm:codice-identificativo-ente isa tmcl:occurrence-t
- 'Codice identificativo dell\'ente'
http://gilgamesh.unipv.it/psi/isaar/#isaar512>
has-datatype(xsd:NMTOKEN);
is -unique().
```

In questo estratto di codice si definiscono il tipo di topic "Agente" (abstract) e le sue sottoclassi "Ente", "Persona", "Famiglia" (sottotipi, As Kind Of di "Agente"). Si indica che i topics di tipo "Agente" dovranno avere almeno un *topic name* di tipo "Forma autorizzata del nome" (definito anche attraverso un *subject identifier*). Infine si indica che i topics di tipo "Ente" potranno avere una o più occorrenze di tipo "Codice identificativo dell'ente" e che queste avranno come tipo di dato NMTOKEN e dovranno contenere valori univoci.

Risultati e casi di utilizzo

Il semantic web è, al momento attuale, ancora una terra promessa da raggiungere attraverso un lungo cammino[12]. Come spesso accade, è piuttosto difficile valutare il grado di maturità di una tecnologia mentre questa sta ancora maturando. L'idea di fondo è quella di sostituire l'attuale web, basato su documenti HyperText Markup Language (HTML) interrelati (con collegamenti esclusivamente monodirezionali), con una rete semantica di classi, con proprie caratteristiche e relazioni (nel linguaggio delle *Topic Maps* parleremmo di soggetti rappresentati da topics con specifiche caratteristiche, come nomi, occorrenze e associazioni). L'utilizzo degli strumenti e delle tecnologie del semantic web potrebbe consentire una maggior integrazione dei dati e delle descrizioni archivistiche e dei servizi collegati, in particolar modo nel caso delle pubblicazioni di fonti in rete. Estendendo la dimostrazione presentata in precedenza a tutti gli elementi degli standard descrittivi archivistici è possibile dimostrare come sia attuabile una gestione delle descrizioni archivistiche direttamente nel web semantico con una sufficiente granularità. Una volta appurato che le descrizioni archivistiche possono essere gestite direttamente nei termini del web semantico, occorre interrogarsi su quali strumenti sia possibile costruire su queste basi e se un simile approccio comporti benefici in termini di gestione, ricerca e presentazione dei dati. Ad esempio un sistema che permetta di gestire relazioni (associazioni) multiple (ognuna valida in un determinato ambito, ad esempio di autorità), garantisce la possibilità di fondere strumenti generalmente irrimediabilmente separati o di operare

[12] Già ora si inizia a parlare di Web 3.0 (Moore 2009), ovvero un ulteriore tassello verso il web semantico ipotizzato da Tim Berners Lee (Berners-Lee, Hendler e Lassila 2001; Shadbolt, Berners-Lee e Hall 2006) e quindi forse da intendere come 0.3 (Grimes 2007; Walsh 2009), ad indicare un passaggio verso un punto di arrivo ancora ben distante.

facilmente ricostruzioni virtuali sulla carta. La possibilità di creare un numero arbitrario di alberi archivistici può essere utilizzata in diversi contesti: dalla possibilità per i ricercatori di riorganizzare la documentazione sulla propria scrivania virtuale di lavoro (*Virtual Desktop Environment*), fino alla possibilità di memorizzare un log dei cambiamenti in un lavoro in progress eventualmente condiviso (permettendo così di ripristinare stadi precedenti di lavoro etc.). Un esempio maggiormente immediato potrebbe essere visto nella possibilità di gestione parallela di campi descrittivi (ad esempio degli ambiti e del contenuto, per utilizzare la casistica prevista dagli standard internazionali) con differenti scope notes dal punto di vista della lingua (per gestire descrizioni multilingue) o dal punto di vista del target (per, eventualmente, fornire una descrizione scolastica o destinata ai ragazzi (Nanni 2005)). Ma i dati strutturati in questa maniera e con una tale granularità di informazione portano ad ulteriori prospettive a cui è necessario dedicare quantomeno un rapido accenno. Strutturare le regole di descrizione sotto forma di vincoli (espressi attraverso il linguaggio vincolato TMCL) permette di costruire facilmente interfacce software flessibili, di facile creazione e modifica: ogni motore *Topic Maps* che gestisca schemi TMCL[13] potrebbe automaticamente generare un'interfaccia di inserimento dati che rispetti i vincoli indicati. In tal senso sarebbe sufficiente cambiare un vincolo per modificare l'interfaccia senza dover operare su codice, base di dati o altro (Vassallo 2010). Infine *Topic Maps* così strutturate potrebbero essere usate come formato neutro di interscambio fra sistemi archivistici; del resto bisogna ricordare come le *Topic Maps* nascano proprio come strumento per fondere indici di manuali tecnici, per cui le procedure di fusione, importazione e esportazione sono essenziali fin dalle origini della

[13] Allo stato dell'arte, l'unico software che offra un supporto a schemi TMCL è Topincs http://www.cerny-online.com/topincs/, sviluppato da Robert Cerny http://www.topicmapslab.de/people/Robert_Cerny.

tecnologia. Nel mondo degli archivi l'esigenza di un formato condiviso e di procedure collaudate di esportazione diventa pressante nel caso della comunicazione dei dati verso i vari aggregatori di banche dati come PLAIN (ora Lombardia Beni Culturali Archivi), SIUSA, SIAS e soprattutto verso il nuovo Portale Archivistico Nazionale all'interno del Sistema Archivistico Nazionale. L'approccio finora proposto per ovviare al problema del flusso dei dati è un approccio rigidamente record centrico che offre, è necessario sottolinearlo, comunque numerosi vantaggi, in particolar modo per ciò che concerne le procedure di verifica e recupero degli aggiornamenti. Le *Topic Maps* invece considerano (o, meglio, possono considerare) il sistema nella sua interezza, inoltre bisogna ricordare come uno dei costrutti fondamentali è proprio l'associazione, pertanto le relazioni sarebbero immagazzinate in maniera distinta, autonoma e sarebbe addirittura possibile importare, ad esempio per fini di aggiornamento, una singola relazione. L'idea essenziale è che non è necessario, in nessun caso e in special modo laddove entrino in gioco la cooperazione e la condivisione delle informazioni, manipolare l'intero sistema espresso in una *topic map*, ma è possibile importare, fondere, riferirsi etc. anche solo a un estratto dell'intero sistema, un frammento che per sua natura può essere facilmente integrato anche dinamicamente, *just in time*. Quest'ultima considerazione apre le porte a un ulteriore ambito di applicabilità delle *Topic Maps*, tipico peraltro delle tecnologie del web semantico, ovvero il *linked data*: un sistema archivistico al proprio interno potrebbe, ad esempio, collegare un complesso archivistico a un soggetto produttore, senza descrivere quest'ultimo né includerlo come *topic*, ma semplicemente indicando nell'associazione un URI o, in ogni caso, un identificativo di soggetto (auspicalmente esposti da un' autorità come il Sistema Archivistico Nazionale). La risoluzione dell'identificativo potrebbe essere proprio un frammento di *topic map* che potrebbe essere letto e

caricato dinamicamente completando il puzzle delle informazioni così costituito.

Conclusioni

In conclusione questi spunti di riflessione sono da intendersi come un privo tentativo di discutere e sviscerare le possibilità di applicare una specifica tecnologia, le *Topic Maps*, nel campo della descrizione archivistica, evidenziandone i vantaggi e le potenzialità. Attraverso l'uso del TMCL, il linguaggio vincolato per esprimere schemi di validazione di *Topic Maps*, si è cercato di dimostrare come sia possibile (in linea teorica e presentando un breve esempio tratto dalle regole previste dallo standard ISAAR(CPF)) esprimere gli elementi previsti dagli standard archivistici in *Topic Maps*. Infine si sono evidenziate le opportunità offerte da un simile approccio in campi affini, ma non limitati alla descrizione, il riordino e l'inventariazione di un complesso archivistico, verificando e proponendo soluzioni da adottare per lo scambio di dati all'interno di sistemi archivistici (con particolare riferimento ad aggregatori come il neonato Portale Archivistico Nazionale all'interno del Sistema Archivistico Nazionale) e per la costruzione di interfacce flessibili e modulari.

Riferimenti bibliografici

AHMED, KAL (2003), «Hierarchical Classification Systems», http://www.techquila.com/tmsinia4.html.

BERNERS-LEE, TIM, JAMES HENDLER e ORA LASSILA (2001), «The Semantic Web», *Scientific American*, 284, 5, pp. 34–43, http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

BRY, FRANÇOIS *et al.* (2006), «Efficient evaluation of n-ary conjunctive queries over trees and graphs», in *WIDM '06: Proceedings of the 8th annual ACM international workshop on Web information and data management*, NewYork: ACM, pp. 11–18, http://rewerse.net/publications/download/REWERSE-RP-2006-130.pdf.

DE GRAAUW, MARC (2002a), «Structuring Scope», http://www.marcdegraauw.com/files/structuring_scope.htm.

— (2002b), «Survey of Actual Scope Use in Topic Maps», http://www.marcdegraauw.com/files/scope_survey.htm.

ECO, UMBERTO (1973), *Il segno*, Milano: Isedi.

GARRIDO, PIEDAD e JESÙS TRAMULLAS (2004), «Topic Maps an alternative or a complement to concept maps?», in *Concept Maps: Theory, Methodology, Technology Proceedings of the First Int. Conference on Concept Mapping, Pamplona 2004*, http://cmc.ihmc.us/papers/cmc2004-226.pdf.

GARSHOL, LARS MARIUS (2001), «Topic maps, RDF, DAML, OIL: a comparison», http://www.ontopia.net/topicmaps/materials/tmrdfoildaml.html.

— (2003), «Living with topic maps and RDF», http://www.ontopia.net/topicmaps/materials/tmrdf.html.

— (2004), «Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all», *Journal of Information Science*, 30, 4, pp. 378–391, http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html.

— (2005), «XTM is not Topic Maps», http://www.garshol.priv.no/blog/7.html.

— (2006), «Ontologies: validation or reasoning», http://www.garshol.priv.no/blog/76.html.

— (2008), «A Theory of Scope», in *Scaling Topic Maps. Third International Conference on Topic Maps Research and Applications, TMRA 2007 Leipzig, Germany, October 11-12, 2007 Revised Selected Papers*, a cura di Lutz Maicher e Lars Marius Garshol, Berlino: Springer-Verlag, pp. 74–85.

GRIMES, SETH (2007), «Tableau does Web 0.2... but that's just a first step», http://intelligententerprise.informationweek.com/blog/archives/2007/11/tableau_does_we.html.

GROENMO, GEIR OVE e STEVE PEPPER (2001), «Towards a General Theory of Scope», in *Proceedings of Extreme Markup Languages 2001*, http://conferences.idealliance.org/extreme/html/2001/Pepper01/EML2001Pepper01.html.

ISO/IEC 13250-2:2006 (2006), *Information Technology - Document Description and Processing Languages - Topic Maps - Data Model*, Geneve: International Organization for Standardization, http://www.isotopicmaps.org/sam/sam-model/.

ISO/IEC CD 29111 (2009), *Information Technology - Topic Maps - Expressing Dublin Core Metadata using Topic Maps*, Geneva: International Organization for Standardization, http://www.jtc1sc34.org/repository/0883.htm.

ISO/IEC FCD 19756 (2010), *Information Technology - Document Description and Processing Languages - Topic Maps - Data Model*, Geneve: International Organization for Standardization, http://www.isotopicmaps.org/tmcl/tmcl.html.

LEE, HYUN-SIL, YANG-SEUNG JEON e SUNG-KOOK HAN (2006), «MARCXTM: Topic Maps Modeling of MARC Bibliographic Information», in *Charting the Topic Maps Research and Applications Landscape, First International Workshop on Topic Maps Research and Applications, TMRA 2005, Leipzig, Germany, October 6-7, 2005, Revised Selected Papers*, a cura di Lutz Maicher e Alexander Sigel, Berlino: Springer-Verlag, pp. 241–252.

MAICHER, LUTZ (2008), «Mapping between the Dublin Core Abstract Model DCAM and the TMDM», in *Scaling Topic Maps, Third International Conference on Topic Maps Research and Applications, TMRA 2007, Leipzig, Germany, October 11-12, 2007, Revised Selected Papers*, a cura di Lutz Maicher e Lars Marius Garshol, Berlino: Springer-Verlag, pp. 198–213, http://www.informatik.uni-leipzig.de/\~maicher/publications/DCAMandTMDM_latest.pdf.

MESCHINI, FEDERICO (2005), «Le mappe topiche: come imparai a non preoccuparmi e ad amare i metadati», *Bollettino AIB*, 45, 1, pp. 59–72, http://www.aib.it/aib/boll/2005/0501059.htm.

MOORE, GRAHAM (2009), «Topic Maps fits in very well with Web 3.0 and Linked Data», http://www.semantic-web.at/1.36.resource.293.graham-moore-x22-topic-maps-fits-in-very-well-with-web-3-0-and-linked-data-x22.htm.

NANNI, GIANCARO (2005), *L'archivio storico spiegato ai ragazzi*, San Miniato: Titivillus, p. 98.

OGDEN, CHARLES KEY e IVOR ARMSTRONG RICHARDS (1923), *The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*, London: Routledge & Kegan Paul.

OH, SAM GYUN (2008), «MARC, FRBR and RDA: The Topic Maps Perspective», in *Topic Maps 2008*, http://www.topicmaps.com/tm2008/oh.pdf.

PEPPER, STEVE (1999), «Navigating haystacks and discovering needles: introducing the new topic map standard», *Markup Language*, 1, 4, pp. 47–74.

— (2000), «The TAO of topic maps: finding the way in the age of infoglut», in *Proceedings of XML Europe*, http://www.sil.org/sgml/pepperTAOofTopicMaps.pdf.

— (2002), «Ten Theses on Topic Maps and RDF», http://www.ontopia.net/topicmaps/materials/rdf.html.

PEPPER, STEVE (2008a), «Expressing Dublin Core in Topic Maps», in *Scaling Topic Maps. Third International Conference on Topic Maps Research and Applications, TMRA 2007 Leipzig, Germany, October 11-12, 2007 Revised Selected Papers*, a cura di Lutz Maicher e Lars Marius Garshol, Berlino: Springer-Verlag, pp. 186–197, http://www.ontopedia.net/pepper/papers/DCinTopicMaps.pdf.

— (2008b), «Topic Maps and the Semantic Web», http://topicmaps.wordpress.com/2008/05/11/topic-maps-and-the-semantic-web/TopicMapsandtheSemanticWeb.

PEPPER, STEVE *et al.* (2006a), «A Survey of RDF/Topic Maps Interoperability Proposals», http://www.w3.org/TR/rdftm-survey/.

PEPPER, STEVE *et al.* (2006b), «Guidelines for RDF/Topic Maps Interoperability», http://www.ontopia.net/work/guidelines.html.

PRESUTTI, VALENTINA *et al.* (2005), «Towards the definition of guidelines for RDF and Topic Maps interoperability», in *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic annotation (SemAnnot 2005) located at the 4rd International Semantic Web Conference (ISWC 2005)*, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-185/semAnnot05-08.pdf.

SHADBOLT, NIGEL, TIM BERNERS-LEE e WENDY HALL (2006), «The Semantic Web Revisited», *IEEE Intelligent Systems*, 21, pp. 96–101, http://doi.ieeecomputersociety.org/10.1109/MIS.2006.62.

SIGEL, ALEXANDER (2003), «Topic Maps in Knowledge Organization», in *XML Topic Maps: creating and using topic maps for the web*, a cura di Jack Park e Sam Hunting, Boston: Addison-Wesley, pp. 383–476.

THOMAS, HENDRIK *et al.* (2008), «GTMalpha - Towards a Graphical Notation for Topic Maps», in *Subject-centric Computing: Fourth International Conference on Topic Maps Research and Applications, TMRA 2008 Leipzig, Germany, October 16-17, 2008. Revised Selected Papers*, a cura di Lutz Maicher e Lars Marius Garshol, Leipzig: Leipzig Univeristy Press, pp. 113–128.

VASSALLO, SALVATORE, in *Workshop Teca del Mediterraneo XI. L'organizzazione della conoscenza fra identità e multiculturalità*, a cura di Maria Abenante.

— (2007), «Le mappe topiche come un ponte tra beni culturali diversi», *Culture del testo e del documento: le discipline del libro nelle biblioteche e negli archivi*, 22, pp. 97–109.

— (2010), «Topincs 4, awesome», http://culturalheritage.wordpress.com/2010/03/30/topincs-4-awesome/.

VASSALLO, SALVATORE e PAUL GABRIELE WESTON (2007), «...e il navigar m'è dolce in questo mare: linee di sviluppo e personalizzazione dei cataloghi», in *La biblioteca su misura: verso la personalizzazione del servizio*, a cura di Claudio Gamba e Maria Laura Trapletti, Milano: Editrice Bibliografica, pp. 130–167.

VIVANET, GIULIANO (2007), «Topic Maps e XTM per l'e-learning», *Journal of e-Learning and Knowledge Society*, 3.

WALSH, DAVID (2009), «Web 0.2 in the Web 2.0 World», http://davidwalsh.name/web-02-web-20-world.

WESTON, PAUL GABRIELE (2002), «Dal controllo bibliografico alle reti documentarie. Il catalogo elettronico nella prospettiva dell'interoperabilità fra sistemi eterogenei», *Biblioteche Oggi*, 5, pp. 44–56.

YEE, MARTHA M. (2009), «Can Bibliographic Data Be Put Directly Onto the Semantic Web?», *Information Technology and Libraries*, 28, 2, pp. 55–80, http://repositories.cdlib.org/postprints/3369/.

Informazioni

L'autore

Salvatore Vassallo
Università degli Studi di Pavia
Email: vassallosalvatore@gmail.com

Web: http://culturalheritage.wordpress.com

Il saggio

Data di submission: 2010-04-20

Data di accettazione: 2010-05-19

Ultima verifica dei link: 2010-05-21

Data di pubblicazione: 2010-06-15